Possibilistic networks parameter learning: Preliminary empirical comparison

Maroua Haddad [1], **Philippe Leray** [1], **Amélie Levray** [2], **Karim Tabia** [2]

1. *LINA-UMR CNRS 6241, University of Nantes, France*
   {maroua.haddad,philippe.Leray}@univ-nantes.fr
2. *Univ Lille Nord de France, F-59000 Lille, France*
   CRIL Univ. Artois CNRS F-62300 Lens, France
   {levray,tabia}@cril.fr

DOI:10.3166/RIA.2016.1-**??**

ike Bayesian networks, possibilistic ones compactly encode joint uncertainty representations over a set of variables. Learning possibilistic networks from data in general and from imperfect or scarce data in particular, has not received enough attention. Indeed, only few works deal with learning the structure and the parameters of a possibilistic network from a dataset. This paper provides a preliminary comparative empirical evaluation of two approaches for learning the parameters of a possibilistic network from empirical data. The first method is a possibilistic approach while the second one first learns imprecise probability measures then transforms them into possibility distributions by means of probability-possibility transformations. The comparative evaluation focuses on learning belief networks on datasets with missing data and scarce datasets.

1. Introduction

Graphical belief models are compact and powerful representations of uncertain information. They allow to factorize a joint uncertainty measure such as a joint probability measure over a set of variables into a set of local measures. This factorization is often based on the conditional independence relations. Belief networks are either built from information elicited directly from experts or learnt automatically by machine learning techniques from empirical data. Possibilistic formalisms are more suitable for representing qualitative and incomplete information. However, there are only few works dealing with learning possibilistic networks from data (Kruse, Borgelt, 1995 ; Borgelt, Kruse, 1998 ; Haddad *et al*., 2015b ; 2015a). In particular, learning a possibilistic network may be sound in case of small datasets or datasets with missing or imprecise information (Dubois, Prade, 2016 ; Haddad *et al*., 2015a ; 2015b).
Learning a graphical belief model comes down in general to i) learn the graphical component, also called structure (namely, extract and encode the independence relationships) and ii) learn the parameters (fill the local tables) associated with each variable. In this paper, we focus on parameter learning of possibilistic networks. Namely, given a structure and a dataset, the goal is to assess local possibility tables of each variable in the context of its parents. The main contributions of the paper are:

– An empirical comparison of two main approaches for learning possibility distributions from data on synthetic datasets. This evaluation compares the networks learnt using two different approaches using a generalized form of the information affinity measure (Jenhani *et al*., 2007).

– An empirical comparison of learning naive possibilistic network classifiers from real datasets. The evaluation here aims to compare the predictive power of possibilistic classifiers learnt from small datasets containing missing data.

2. Graphical belief models

Belief graphical models such as Bayesian (Darwiche, 2009), credal (Cozman, 2000) and possibilistic networks (C. Borgelt, Kruse, 2003) are powerful tools for representing and reasoning with uncertain information. Bayesian networks allow to compactly encode a probability distribution thanks to the conditional independence relationships existing between the variables. Credal networks, based on the theory of credal sets, generalize Bayesian networks in order to allow some flexibility regarding the model parameters. They are for instance used in robustness analysis and for encoding incomplete and ill-known knowledge and reasoning with the knowledge of groups of experts. Possibilistic networks are the counterparts of Bayesian networks based on possibility theory (Dubois, Prade, 1988 ; Zadeh, 1999), more suited for handling imperfect, qualitative and partial information.

2.1. Bayesian networks

Bayesian networks ($\mathcal{BN}$) are well-known probabilistic graphical models

(Darwiche, 2009). They allow to compactly represent a probability distribution over a set of variables of interest. A $\mathcal{BN}$=<G,Θ> is specified by:

– *A graphical component* G with nodes and arcs forming a directed acyclic graph (DAG). Each node represents a variable A_i of the modeled problem and the edges encode independence relationships among variables.

– *A quantitative component* Θ, where each variable A_i is associated with a local probability distribution $p(A_i|par(A_i))$ for A_i in the context of its parents $par(A_i)$.

The joint probability distribution encoded by a Bayesian network is computed using the following chain rule:

$$P(A_1, .., A_n) = \prod_{i=1}^{n} P(A_i|par(A_i)) \quad (1)$$

2.2. Credal networks

Credal networks are probabilistic graphical models based on imprecise probabilities. These latter could be encoded using different representations such as the vertex-based representation (where the credal set is convex and it is characterized by its extreme points), the interval-based representation (where each event is associated with an interval of probability degrees), etc. A credal network $\mathcal{CN}$=<G,K> is a probabilistic graphical model where

– G=<V, E> is a directed acyclic graph (DAG) encoding conditional independence relationships where V=$\{A_1, A_2, .., A_n\}$ is the set of variables of interest (D_i denotes the domain of variable A_i) and E is the set of arcs of G.

– K=$\{K_1, K_2, .., K_n\}$ is a collection of local credal sets, each K_i is associated with the variable A_i in the context of its parents $par(A_i)$.

Such credal networks are called separately specified credal networks as the only constraints on probabilities are specified in local tables for each variable in the context of its parents. Note that in practice, in local tables, one can either specify a set of extreme points characterizing the credal set as in JavaBayes [1] software or directly local interval-based probability measures.
A credal network $\mathcal{CN}$ can be seen as a set of Bayesian networks $\mathcal{BN}$s, each encoding a joint probability distribution. One can compute an interval-based joint probability distribution as follows:

$$\underline{P}(a_1 a_2..a_n) = \min_{p \in ext(\mathcal{CN})} (p(a_1 a_2..a_n)) \quad (2)$$

$$\overline{P}(a_1 a_2..a_n) = \max_{p \in ext(\mathcal{CN})} (p(a_1 a_2..a_n)) \quad (3)$$

In Equations 2 and 3, $p(a_1 a_2..a_n)$ is computed with the chain rule (see Equation 1). Note that the vertices of $ext(\mathcal{CN})$ can be obtained by considering only the set of vertices of the local credal sets K_i associated with the variables (Cozman, 2000).

1. http://www.cs.cmu.edu/~javabayes/Home/

2.3. Possibilistic networks

A possibilistic network $\mathcal{PN}$=<G,Θ> is specified by:

i) *A graphical component* G consisting of a directed acyclic graph (DAG) where vertices represent the variables and edges represent direct *dependence* relationships between variables.

ii) *A numerical component* Θ allowing to weight the uncertainty relative to each variable using local possibility tables. The possibilistic component consists in a set of local possibility tables θ_i=$\pi(A_i|par(A_i))$ for each variable A_i in the context of its parents $par(A_i)$ in the network $\mathcal{PN}$.

Note that all the local possibility distributions θ_i must be normalized, namely $\forall i$=1..n, for each parent context $par(a_i)$, $\max_{a_i \in D_i}(\pi(a_i \mid par(a_i))$=1. See in Figure 1 an example of a Bayesian network and an example of a possibilistic one.
In the possibilistic setting, the joint possibility distribution is factorized using the following possibilistic counterpart of the chain rule:

$$\pi(a_1, a_2, .., a_n) = \otimes_{i=1}^{n}(\pi(a_i|par(a_i))). \quad (4)$$

where $\otimes$ denotes the product or the min-based operator depending on the quantitative or the qualitative interpretation of the possibilistic scale (Dubois, Prade, 1988). In this work, we are interested only in product-based possibilistic networks since we view possibility degrees as upper bounds of probability degrees.

3. Learning the parameters of a possibilistic network

Learning the parameters of a possibilistic network is the problem of assessing the entries of local possibility tables $\pi(A_i|par(A_i))$ for each variable A_i given a structure $\mathcal{S}$ and a dataset $\mathcal{D}$. The structure here is assumed to be given (eg. when learning naive classifiers, the structure is fixed in advance by assumption) or learnt automatically. There are basically two ways to learn the parameters (Haddad *et al.*, 2015a ; 2015b): i) Transformation-based approach (TA for short) and ii) Possibilistic-based approach (PA for short). Note that the authors in (Serrurier, Prade, 2015) propose a possibilistic-based method for learning the structure of a Bayesian network.

3.1. Transformation-based approach

This approach consists in first learning the parameters of a probabilistic network then transforming the obtained probabilistic network into a possibilistic one (Haddad *et al.*, 2015a ; 2015b ; Benferhat *et al.*, 2015 ; Slimen *et al.*, 2013). Transformations can be useful in various contexts such as using the existing tools (e.g. algorithms and software) developed in one setting rather than developing everything from scratch for the other setting or exploiting information provided in different uncertainty languages as it is often the case in some multiple expert applications.

Many probability-possibility transformations exist (Dubois *et al*., 2004 ; Klir, Geer, 1993 ; Zadeh, 1999) but most of the works deal with the desirable properties and propose transformations satisfying such properties. Among these transformations, the optimal transformation (OT) (Dubois *et al*., 2004) is defined as follows:

$$\pi_i = \sum_{j/p_j \leq p_i} p_j, \tag{5}$$

where π_i (resp. p_i) denotes $\pi(\omega_i)$ (resp. $p(\omega_i)$). The transformation of Equation 5 transforms p into π and guarantees that the obtained possibility distribution π is the most specific [2] (hence most informative) one that is consistent and preserving the order of interpretations.

In case where the probabilistic model is a credal one (Bayesian networks where local tables encode sets of probability measures), one can make use of imprecise probability - possibility transformations turning for instance an interval-based probability distribution (IPD) into a possibilistic one. For instance, the transformation proposed in (Masson, Denoeux, 2006) allows to find a possibility distribution dominating all the probability measures defined by probability intervals. This transformation tries on the one hand to preserve the order of interpretations induced by the IPD and the dominance principle requiring that $\forall\phi\subseteq\Omega$, $P(\phi)\leq\Pi(\phi)$ on the other hand. Such transformations correspond to viewing possibility degrees as upper bounds of probability degrees (Dubois *et al*., 1993). There are two main drawbacks with the transformation proposed in (Masson, Denoeux, 2006) : i) its computational cost (it considers in the worst case $N!$ linear extensions where N is the size of the distribution to transform) and ii) the fact that this transformation does not guarantee that the obtained distribution is optimal in terms of specificity. Indeed, it was shown in (Destercke *et al*., 2007) that the transformation of (Masson, Denoeux, 2006) results in a loss of information as it is not the most specific one dominating the considered interval-based probability distribution.
In (Destercke *et al*., 2007), the authors claim that any upper generalized R-cumulative distribution $\overline{F}$ built from one linear extension can be viewed as a possibility distribution and it also dominates all the probability distributions that are compatible with the IPD. Let $\mathcal{C}_l$ be a linear extension compatible with the partial order $\mathcal{M}$ induced by an IPD. Let ϕ_1, $\phi_2..\phi_n$ be subsets of Ω such that ϕ_i=$\{\omega_j|\omega_j \leq_{\mathcal{C}_l} \omega_i\}$. The upper cumulative distribution $\overline{F}$ built from one linear extension $\mathcal{C}_l$ is as follows (see (Destercke *et al*., 2007) for more details):

$$\overline{F}(\phi_i) = \min(\sum_{\omega_j \in \phi_i} u_j, 1 - \sum_{\omega_j \notin \phi_i} l_j) \tag{6}$$

The obtained cumulative distribution $\overline{F}$ is a possibility distribution dominating the IPD and it is such that $\overline{P}(\phi_i)$=$\Pi(A_i)$. The advantage of such a transformation, also called p-box transformation, is its low computational cost (linear in the size of domains) and the fact that the obtained distribution is better in terms of specificity (meaning that the transformation process losses less information).

2. Let π' and π'' be two possibility distributions, π' is more specific than π'' iff $\forall\omega_i\in\Omega$, $\pi'(\omega_i)\leq\pi''(\omega_i)$

Now, using a probability-possibility transformation, one can turn a Bayesian or credal network $\mathcal{BN}$ into a possibilistic network $\mathcal{PN}$ only by turning every local probability table into a possibilistic one while preserving the structure of the network in order to preserve the conditional independences encoded by $\mathcal{BN}$.

EXEMPLE 1. — Let $\mathcal{BN}$ be the Bayesian network of Figure 1 over two variables A and B having the domains D_A=$\{a_1, a_2\}$ and D_B=$\{b_1, b_2, b_3\}$ respectively.

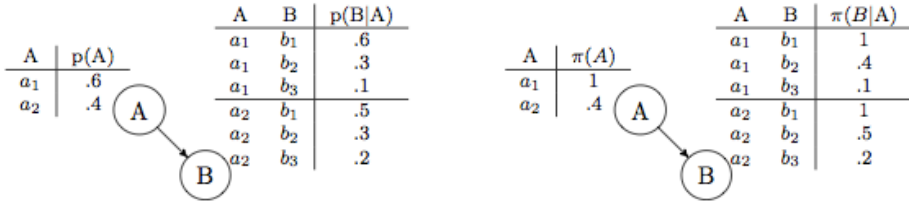

FIGURE *1. Example of a Bayesian network $\mathcal{BN}$ and the possibilistic network $\mathcal{PN}$ obtained from $\mathcal{BN}$ using the optimal transformation OT.*

□

3.2. Possibilistic-based approach

One view of possibility theory is to consider a possibility distribution π on a variable A_i as a *counter function* of a random set (Shafer *et al.*, 1976) pertaining to D_i, the domain of A_i. A random set in D_i is a random variable which takes its values on subsets of D_i. More formally, let D_i be a finite domain. A basic probability assignment or mass function is a mapping $m : 2^{D_i} \longmapsto [0, 1]$ such that $\sum_{a_i \subseteq D_i}(m(a_i))$=1 and $m(\emptyset)$=0. A set $a_i \subseteq D_i$ such that $m(a_i)>0$ is called a focal set.

The possibility degree of an event a_i is the probability of the possibility of the event i.e. the probability of the disjunction of all events (focal sets) a'_i in which this event is included (C. Borgelt *et al.*, 2009):

$$\pi(a_i) = \sum_{a'_i | a_i \cap a'_i \neq \emptyset} m(a'_i) \tag{7}$$

A random set is said to be *consistent* if there is at least one element a_i contained in all focal sets a'_i and the possibility distribution induced by a consistent random set is, thereby, normalized. Exploring this link between possibility theory and random sets theory has been extensively studied, in particular, in learning tasks, we cite for instance (C. Borgelt *et al.*, 2009 ; Joslyn, 1997). In what follows, we investigate this link to learn possibilistic networks parameters from imprecise data.

Given a DAG and an imprecision degree S_i, let $\mathcal{D}_{ij} = \{d_{ij}^{(l)}\}$ be a dataset relative to a variable A_i, $d_{ij}^{(l)} \in D_{ij}$ (resp. $d_{ij}^{(l)} \subseteq D_{ij}$) if data are precise (resp. imprecise). The number of occurrences $A_i = a_{ik}$ such that $Pa(A_i) = j$, denoted

by N_{ijk}, is the number of times $A_i = a_{ik}$ such that $Pa(A_i) = j$ appears in $\mathcal{D}_{ij}$: $N_{ijk} = \text{card}(\{l \text{ s.t. } A_i = a_{ik} \text{s.t.} Pa(A_i) = j \in d_{ij}^{(l)}\})$.

$$\pi(A = a_{ik} | \hat{Pa}(A_i) = j) = \frac{N_{ijk}}{\sum_{k=1}^{r_i} N_{ijk}} * S_i \tag{8}$$

where q_i is $\text{card}(Pa(X_i))$, $r_i = \text{card}(D_i)$ and S_i corresponds to the imprecision degree relative to a variable A_i. To obtain normalized possibility distributions, we divide each obtained distribution by its maximum. It is evident that this operation eliminates S_i. However, we could assign to each value of X_i an imprecision degree which could be either set by an expert or inferred from the dataset to learn from.

4. Experimental study 1: How similar are possibilistic networks where the parameters are learnt using different approaches?

This section provides experimental results comparing the two approaches for learning the parameters of a possibilistic network from *imprecise* data.

4.1. Experimentation setup

In this experiment, given a dataset $\mathcal{D}$ and a network structure (DAG) $\mathcal{S}$, we compare learning a possibilistic network parameters using two approaches, TA and PA. We denote by G^{TA} (resp. G^{PA}) the possibilistic network having the structure $\mathcal{S}$ and its parameters are learnt over the dataset $\mathcal{D}$ using the transformation-based approach TA based on the p-box transformation (resp. the possibilistic-based approach PA).
When learning belief graphical models like Bayesian or possibilistic networks from data, the evaluation is generally carried out by comparing reference networks with the learnt ones. Reference networks are graphical models that are either chosen by an expert or randomly generated. From the reference model, a dataset is generated following the distribution encoded by the reference model. This dataset is then used to learn models using the approach to be evaluated. The problem then comes down to compare the learnt model with the reference one. A comparison may take into account only the joint measures encoded by the learnt and the reference models. In addition, one may want also to take into account the structure of the learnt and reference models. Given that we are only interested in comparing possibilistic networks with same structure, there is no need to consider the graphical component in our comparisons. One simple but costly way of comparing the reference network with the learnt one is to compare only the joint distribution encoded by the reference model with the learnt model distribution. An example of similarity measure for possibility distributions is information affinity (Jenhani *et al.*, 2007). However the size of the distribution may be very huge (it fact, it is exponential in the number of variables of the network) making it impossible to compare joint possibility distributions. We propose a heuristic method that compares the networks' local distributions locally and aggregates the results to

provide an overall similarity score of two possibilistic networks.
Information affinity (Jenhani *et al*., 2007) is defined as follows:

$$InfoAff(\pi_1, \pi_2) = 1 - \frac{d(\pi_1, \pi_2) + Inc(\pi_1, \pi_2)}{2} \quad (9)$$

where $d(\pi_1, \pi_2)$ represents the mean Manhattan distance between possibility distributions π_1 and π_2 and defined as follows: $d(\pi_1, \pi_2)$=$\frac{1}{N}\sum_{i=1}^{N} |\pi_1(\omega_i) - \pi_2(\omega_i)|$. As for $Inc(\pi_1, \pi_2)$, it is a measure of inconsistency and it assesses the conflict degree between π_1 and π_2. Namely, $Inc(\pi_1, \pi_2)$=1-$\max_{\omega_i \in \Omega}(\pi_1(\omega_i) \wedge \pi_2(\omega_i))$ where $\pi_1(\omega_i) \wedge \pi_2(\omega_i)$ denotes a combination operation of two possibility distributions. In (Jenhani *et al*., 2007) , the $\min$ operator is used in a qualitative setting. In a quantitative setting, a product operator can be used as well.

EXEMPLE 2. — Let Ω=$\{\omega_1, \omega_2, \omega_3, \omega_4\}$ and let π_1 and π_2 be two possibility distributions such that π_1=$(1, .8, .4, 0)$ and π_2=$(.9, 1, .2, .2)$. Then $InfoAff(\pi_1, \pi_2)$=1-$\frac{d(\pi_1,\pi_2)+Inc(\pi_1,\pi_2)}{2}$=1-$\frac{d(0,175+0,1)}{2}$ = $.8625$. □

The measure of Equation 9 satisfies many natural properties such as *non-negativity, symmetry, upper bound and non-degeneracy, lower bound, inclusion* and *permutation*. However, it cannot be directly applied for assessing the similarity of two possibilistic networks. Now, in order to assess the similarity between two possibilistic networks G_1 and G_2 having the same structure (same DAG), then it may be relevant to compare locally every local possibility distribution π_1^i in the network G_1 with π_2^i, namely its corresponding distribution in G_2. This can be done for instance using an aggregation function (in this work, we use the mean) that takes into account all the local distributions and returns a global similarity score between $G1$ and $G2$.

$$GrInfoAff(G_1, G_2) = \frac{1}{N}(\sum_{i=1..N} InfoAff(\pi_1^i, \pi_2^i)), \quad (10)$$

where N stands for the number of local distributions of the networks.

4.2. Benchmarks

We first generated a set of possibilistic networks with different features (number of variables, number of parents per variable, rate of imprecise data, etc.). For each possibilistic network G, we generate datasets according to G. More precisely, for each possibilistic network G (characterized by its number of variables denoted $\#$ *variables*, the mean number of parents per node denoted μ *variables* and the mean domain size of variables μ *domain*), we generate many datasets (with different sizes). Regarding the dataset generation process, it consists in generating an imprecise dataset representative of its possibility distribution. The sampling process constructs a database of N (predefined) observations by instantiating all variables w.r.t. their possibility distributions using the α-*cut* notion expressed as follows:

$$\alpha - \text{cut}_{A_i} = \{a_i \in D_i \text{ s.t. } \pi(a_i) \geq \alpha\} \quad (11)$$

where α is randomly generated from [0,1].

Obviously, variables are most easily processed w.r.t. a topological order, since this ensures that all parents are instantiated. Instantiating a parentless variable corresponds to computing its α-cut. Instantiating a conditioned variable A_i s.t. $Pa(A_i = A)$ corresponds to computing the α-cut of $\pi(A_i|Pa(A_i) = A)$ computed as follows:

$$\pi(A_i|Pa(A_i) = A) = \max_{a_i \in A}(\pi(A_i|a_i), \pi(a_i)) \quad (12)$$

Table 1 gives the details on the generated possibilistic networks and the corresponding datasets.

Tableau 1. Datasets properties used in experiments 1.

Name	# variables	μ parents	μ domain	# datasets
Net10	10	1.6	3.9	9
Net20	20	2.65	3.41	8
Net30	30	2.76	3.48	7

4.3. Results

Table 2 gives the results of computing the similarity on each dataset $\mathcal{D}_i$, the possibilistic network G_i^{TA} (resp. G_i^{PA}) learnt using the TA (resp. PA) approach with the reference network G_i used to generate D_i.

Tableau 2. Results of experiments 1.

Dataset	TA	PA
Net10	0.63	0.8
Net20	0.64	0.8
Net30	0.67	0.8

The results of Table 2 show that on the one hand the learnt possibilistic networks using the TA approach are close to the reference ones. Namely, they have rather a good similarity with the reference possibilistic networks used to generate the datasets. Moreover, the obtained similarity scores do not seem to be affected by the number of variables, variable domains size, etc. Regarding the possibilistic networks learnt using the PA approach, their similarity scores are slightly better, but this is expected as the datasets generation process and the PA approach have the same view of possibility degrees. Such results also rise the issue of similarity measures on possibilistic networks which is still an open issue.

5. Experimental study 2: Predictive power of possibilistic network classifiers where the parameters are learnt using different approaches

In this section, we evaluate the predictive power of credal network classifiers (Corani, Zaffalon, 2008) with respect to their possibilistic network counterparts. More

precisely, we compare on many datasets the classification efficiency of naive credal classifier (NCC for short) and the corresponding possibilistic classifiers obtained either using the possibilistic-based approach (PNC_{PA}) or using the transformation-based approach (PNC_{TA}). Moreover, we compare our results to naive Bayes classifier (NBC) as a baseline.

5.1. Experimentation setup

In order to evaluate the NCC classifier, we use the following measures used in (Corani, Zaffalon, 2008).

– *Determinacy (Det):* It is the percentage of predictions outputting a unique (precise) class label.

– *Single-Accuracy (SiAcc):* It denotes the percentage of correct classifications when the predictions of NCC are precise.

– *Set-Accuracy (SetAcc)*: It is the proportion of imprecise predictions containing the right class label.

The evaluation mode used in this experiment is a 10-fold cross validation.

5.1.1. Benchmarks

The experimental study is carried out on the following datasets where some data values are missing. Note the missing data is assumed to be not missing at random. The first four datasets of Table 3 are real datasets used in the literature for evaluating classifiers with missing data (http://sci2s.ugr.es/keel/missing.php). The remaining ones are collected from different sources.

Tableau 3. Datasets used in our experiments.

Name	# instances	# variables	# classes	% missing
breast	286	9	2	4 %
housevotes	435	16	2	24 %
mushroom	8124	22	2	31 %
post-operative	90	8	3	3 %
audiology	226	70	24	98%
sick	3772	30	2	20%
primary-tumor	339	18	21	46%
kr-vs-kp	3196	37	2	0 %
soybean	683	36	19	18%
crx	690	16	2	2%

5.1.2. Results

Table 4 gives the results of evaluating the NCC classifier on the datasets of Table 3.

Tableau 4. Results of the NCC classifier on the datasets of Table 3.

Dataset	Det	SiAcc	SetAcc
breast	92.43 %	74.08 %	100 %
housevotes	99.52 %	90.26 %	100 %
mushroom	96.10 %	99.56 %	100 %
post-operative	49.67 %	67.57 %	84.36 %
audiology	7.76%	99.55%	99.03%
sick	98.93%	97.54%	100%
primary-tumor	13.59%	77.11%	63.37%
kr-vs-kp	99.18%	88.16%	100%
soybean	47.38%	92.56%	97.85%
crx	94.01%	86.34%	100%

The results of Table 4 show good single accuracy rates with high determinacy rates except for the $post\text{-}operative$, $audiology$, $primary\text{-}tumor$ and $soybean$ datasets. Typically, it's on small datasets with many classes where the NCC is not efficient.

Table 5 gives the results of evaluating the NBC (Naive Bayes Classifier), PNC_{TA} and PNC_{PA} classifiers on the datasets of Table 3.

Tableau 5. Results of the NBC, PNC_{PA} and PNC_{TA} classifiers on the datasets of Table 3.

	% of correct classifications		
Dataset	NBC	PNC_{PA}	PNC_{TA}
breast	**72.88%**	72.73 %	70.27%
housevotes	**90.11 %**	89.19 %	58.71 %
mushroom	**95.73 %**	77.35 %	85.34 %
post-operative	68.11 %	67.78 %	**71.11%**
audiology	**72.79%**	55.90%	11.54%
sick	**96.97%**	95.53%	94.41%
primary-tumor	**49.54%**	28.42%	43.42%
kr-vs-kp	**87.82%**	85.86%	86.89%
soybean	92.66%		75.51 %
crx	85.38%	85.80%	**91.01%**

Results of Table 5 show that classifiers NBC, PNC_{PA} and PNC_{TA} have most of the time comparable results in terms of correct classification rates on some datases but they show real performances on some other datasets. This is also valid for the results of the NCC classifier. Now, comparing PNC_{PA} and PNC_{TA}, this latter achieves better results on two datasets while the former has better classification rates on the two other datasets. It is not obvious what makes a given approach better, a thourough analysis of the properties of the datasets is needed to help understanding such results.

6. Discussions and concluding remarks

This paper provides preliminary results comparing two methods for assessing the parameters of a possibilistic network given a structure and a dataset. The first experiment shows that the possibilistic-based method learns better and more information in terms of information affinity than the method based on the probability-possibility transformation. This is not surprising since the data was generated according to the possibility distributions of the reference networks. This also confirms that there is some information loss when transforming probability distributions into possibilistic ones. Another important result is that the classifiers based on possibilistic networks have somehoww comparable efficiency with naive Bayes and credal classifiers. On the other hand, the possibilistic classifiers where the parameters have been learned with two different approaches have basically comparable results. Overall, these results show that no system really outperforms the others on all the datasets. Such results are preliminary but encouraging, a further comparative study on a large number of benchmarks and problems (classification and inference in general), will be needed to really compare the two approaches. The question that also arises is how to know for a given or a particular dataset which kind of classifier to use?

Acknowledgements

This work is done with the support of a CNRS funded project PEPS FaSciDo 2015 - MAPPOS.

Bibliographie

Benferhat S., Levray A., Tabia K. (2015). On the analysis of probability-possibility transformations: Changing operations and graphical models. In *ECSQARU 2015, compiegne, france, july 15-17, 2015. proceedings.*

Borgelt, Kruse. (1998). Probabilistic and possibilistic networks and how to learn them from data. In, p. 403–426. Springer.

Borgelt C., Kruse R. (2003, Apr). Learning possibilistic graphical models from data. *Fuzzy Systems, IEEE Transactions on*, vol. 11, n° 2, p. 159-172.

Borgelt C., Steinbrecher M., Kruse R. R. (2009). *Graphical models: representations for learning, reasoning and data mining* (vol. 704). Wiley.

Corani G., Zaffalon M. (2008, avril). Learning reliable classifiers from small or incomplete data sets: The naive credal classifier 2. *Journal of Machine Learning Research*, vol. 9, p. 581–621. Consulté sur http://www.jmlr.org/papers/volume9/corani08a/corani08a.pdf

Cozman F. G. (2000). Credal networks. *Artificial Intelligence*, vol. 120, n° 2, p. 199 - 233. Consulté sur http://www.sciencedirect.com/science/article/pii/S0004370200000291

Darwiche A. (2009). *Modeling and reasoning with bayesian networks*. Cambridge University Press.

Destercke S., Dubois D., Chojnacki E. (2007). Transforming probability intervals into other uncertainty models. In *EUSFLAT 2007 proceedings*, vol. 2, p. 367-373. Ostrava, Czech Republic, Universitas Ostraviensis.

Dubois D., Foulloy L., Mauris G., Prade H. (2004). Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, vol. 10, n° 4, p. 273-297.

Dubois D., Prade H. (1988). *Possibility theory: An approach to computerized processing of uncertainty*. New York, Plenum Press.

Dubois D., Prade H. (2016). Practical methods for constructing possibility distributions. *Int. J. Intell. Syst.*, vol. 31, n° 3, p. 215–239.

Dubois D., Prade H., Sandri S. (1993). On possibility/probability transformations. In R. Lowen, M. Roubens (Eds.), *Fuzzy logic*, p. 103-112. Dordrecht, Kluwer Academic Publishers.

Haddad M., Leray P., Amor N. B. (2015a). Evaluating product-based possibilistic networks learning algorithms. In *Symbolic and quantitative approaches to reasoning with uncertainty - 13th european conference, ECSQARU 2015, compiègne, france, july 15-17, 2015. proceedings*, p. 312–321.

Haddad M., Leray P., Amor N. B. (2015b). Learning possibilistic networks from data: a survey. In *2015 conference of the international fuzzy systems association and the european society for fuzzy logic and technology (ifsa-eusflat-15), gijón, spain., june 30, 2015.*

Jenhani I., Amor N. B., Elouedi Z., Benferhat S., Mellouli K. (2007). Information affinity: A new similarity measure for possibilistic uncertain information. In K. Mellouli (Ed.), *Ecsqaru*, vol. 4724, p. 840-852. Springer.

Joslyn C. (1997). Measurement of possibilistic histograms from interval data. *International Journal Of General System*, vol. 26, n° 1-2, p. 9–33.

Klir G. J., Geer J. F. (1993). Information-preserving probability- possibility transformations: Recent developments. In R. Lowen, M. Roubens (Eds.), *Fuzzy logic*, p. 417-428. Dordrecht, Kluwer Academic Publishers.

Kruse R., Borgelt C. (1995). Learning probabilistic and possibilistic networks : Theory and applications. In *Eighth european conference on machine learning*, p. 3–16.

Masson M.-H., Denoeux T. (2006). Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, vol. 157, n° 3, p. 319-340.

Serrurier M., Prade H. (2015). Learning structure of bayesian networks by using possibilistic upper entropy. In P. Grzegorzewski, M. Gagolewski, O. Hryniewicz, Á. M. Gil (Eds.), *Strengthening links between data analysis and soft computing*, p. 87–95. Cham, Springer International Publishing. Consulté sur http://dx.doi.org/10.1007/978-3-319-10765-3_11

Shafer G. *et al*. (1976). *A mathematical theory of evidence* (vol. 1). Princeton university press Princeton.

Slimen Y. B., Ayachi R., Amor N. B. (2013). Probability-possibility transformation: - application to bayesian and possibilistic networks. In F. Masulli, G. Pasi, R. R. Yager (Eds.), *Wilf*, vol. 8256, p. 122-130. Springer.

Zadeh L. A. (1999). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, vol. 100, p. 9–34.